AI DRIVEN HAND GESTURE RECOGNITION

A COMMUNITY SERVICE PROJECT REPORT

Submitted by

Marisetty Sai Murali (9919004342)
Vedala Sai Seetha Ramakrishnamacharyulu (9919004291)
Yalla Sireesha (9919004304)
Yalla Satya Praveen kumar Reddy (9919004303)

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION

KRISHNANKOIL 626 126

April 2022

Major Design Experience Information

Student Details	:	1.Marisetty Sai Murali (9919004342) 2.Vedala Sai Seetha RamaKrishnamacharyulu (9919004291) 3.Yalla Sireesha (9919004304) 4.Yalla Satya Praveen Kumar Reddy (9919004303)
Project Supervisor	:	Mr.R.Raja Subramanian
Project Title	:	AI DRIVEN HAND GESTURE RECOGNITION
Program Concentration Area	:	Sign Language Recognition for Deaf and Hard of hearing students.
Subject(s) as Pre-requisite	:	1. CSE18R173 – Design and Analysis of Algorithms 2. CSE18R257 – Predictive Analytics 3. CSE18R212 – Machine Learning 4. CSE18R396 – Deep Learning 5. CSE18R254 – Introduction to Python Programming
Constraints	:	The proposed Ai-driven gesture recognition system provides reliable results for predicting the hand gestures leveraging Deep Learning algorithms. The System still has constraints with varied gestures and different environment.
Project Related to	:	Human Computer Interaction, Deaf-mute People.
Standard	:	IEEE P2841 – IEEE Draft Framework and Process for Deep Learning Evaluation

DECLARATION

I hereby certify that the work which is being presented in the B.Tech. Community Service Project Report entitled **"AI Driven Hand Gesture Recognition",** in partial fulfillment of the requirements for the award of the **Bachelor of Technology in Computer Science and Engineering** and submitted to the Department of **Computer Science and Engineering** of Kalasalingam Academy of Research and Education (Deemed to be University) – Tamil Nadu., is an authentic record of my own work carried out during a period from December 2021 to April 2022 under the supervision of **Mr. R. Raja Subramanian**.

The matter presented in this thesis has not been submitted by me for the award of any other degree elsewhere.

Signatureof Candidate

M SAI MURALI (9919004342)

V RAMA KRISHNA (9919004291)

Y SIREESHA (9919004304)

Y S PRAVEEN KUMAR (991004303)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date:

Signature of supervisor

Mr. R. Raja Subramanian
Assistant Professor
Computer Science and Engineering

Head

Department of Computer Science and Engineering

Kalasalingam Academy of Research and Education (Deemed to be University)

KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report "**AI DRIVEN HAND GESTURE RECOGNITION"** is the bonafide work of **"M SAI MURALI (9919004343) , V RAMAKRISHNA (9919004291) , Y SIREESHA (9919004304) , Y S PRAVEEN KUMAR (9919004303)"** who carried out the project work under my supervision.

Signature	Signature
Dr. P.DEEPA LAKSHMI	**Mr.R.RAJA SUBRAMANIAN,**
Head of the Department	**Project Supervisor**
Associate Professor	**Assistant Professor**
Department of CSE	**Department of CSE**
Kalasalingam Academy of	Kalasalingam Academy of
Research and Education	Research and Education
Krishnankoil-626126	Krishnankoil-626126
Virudhunagar District.	Virudhunagar District.

Project Viva-voce held on___________

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

First and foremost, we thank the ‘Supreme Power’ for the immense grace showered on us which enabled us to do this project. We take this opportunity to express by sincere thanks to the late, “Kalvivallal” Thiru T. Kalasalingam, Chairman, Kalasalingam Group of Institutions, “Illayavallal” Dr. K. Sridharan,Ph.D., Chancellor, Dr. S. Shasi Anand, Ph.D., Vice President, who is the guiding light for all the activities in our University.

We thank our Vice chancellor Dr.R. Nagaraj, Ph.D., for guiding every one of us and infusing us the strength and enthusiasm to work over successful.

We wish to express our sincere thanks to our respected Head of the Department Dr.P.DEEPA LAKSHMI, whose moral support encouraged us to process through our project work successfully.

We offer our sincerest gratitude to our Supervisor, Mr. R. RAJA SUBRAMANIAN, for his patience, motivation, enthusiasm and immense knowledge.

We are extremely grateful to our Overall Project Coordinator, Dr.T.DHILIPHAN RAJKUMAR for constant encouragement in the completion of the Community Service Project.

Finally, we thank all, our Parents, Faculty, Non-Teaching Faculty and our friends for their moral support.

ABSTRACT

Human–Computer Interfaces (HCI) is the study of how humans and computers interact. Hand gestures are a great way to communicate with individuals who don't comprehend what we're saying, especially when they don't understand what we're saying. It is also an essential component of human–computer interaction. Understanding hand gestures is essential for ensuring that listeners comprehend what speakers are saying or that robots grasp what humans are saying. The main goal of this project is to experiment with various techniques to Hand Gesture Recognition.In this study, we initially used radar data and subsequently camera sensors to perform hand gesture identification. We first attempted to construct hand gesture recognition using radar data, but because most people do not know sign language and interpreters are few, we built a real-time technique for fingerspelling-based American sign language using neural networks, followed by another model using MediaPipe. We present a convolutional neural network (CNN) technique for recognising human hand motions from camera pictures. The goal is to distinguish human work activity hand motions from camera photos. To generate CNN training and test data, skin models, hand positions, and orientations are used. The hand is first sent through the filter, and then it is passed through a classifier that predicts the class of hand motions. The purpose of the hand position is to convert and rotate the picture of the hand into a neutral attitude. The calibrated picture is then used to train CNN. To create this model, we combined Machine Learning, Deep Learning, and computer vision. For detecting various gestures, our MediaPipe Model performs admirably.

TABLE OF CONTENTS

LIST OF FIGURES

Table 1: List of Figures

LIST OF ABBREVIATIONS

Abbreviation	Full form
CNN	Convolutional Neural Network
ANN	Artificial Neural Networks
RNN	Recurrent Neural Networks
LSTM	Long Short Term Memory

Table 2: List of Abbreviations

CHAPTER I

INTRODUCTION

Over the previous few decades, mobile and computing technologies have advanced dramatically. In addition, the way people engage with machines has improved. Human–computer interactions (HCI) approaches range from simple keyboard inputs to advanced vision-based gesture detection systems. Hand gesture recognition is one of the most exciting and important HCI technologies. Hand gesture recognition is an intriguing research subject since it may be used to improve communication in a number of applications such as mobile phones, smart TVs, robot controllers, medical equipment, access control systems, and smart automobiles.

Few Applications are:

- On Air Interaction with Computer
- Medical Field
- Gesture Based Gaming Control
- Autonomous Navigation System
- Controlling Smart Gadgets
- Communication

1.1 Overview

Overview Over the last decade, there has been an increasing interest in developing HCI based on Hand Gesture Recognition employing radar and other RF sensors (HGR). RADAR can penetrate clouds, fogs, mist, and insulators. It is capable of determining the precise position of an item. It can compute the velocity of a target.

It can calculate the distance between two objects. Radar has numerous benefits over other sensors, including the ability to discern between fixed and moving objects. As a result, we will be utilizing radar acquired data for this project.

People with speech impairments must rely on sign language since they are unable to communicate using their hearing or voice. Everyone who is speech handicapped uses sign language, but they have a difficult time interacting with non-signers (those who aren't fluent in sign language). For persons who are deaf or hard of hearing, a sign language interpreter is an essential. Their informal and formal communication is hampered as a result of this Recent advancements in deep learning have resulted in significant improvement in the fields of gesture recognition and motion recognition.

In real time, the proposed technique attempts to convert hand gestures into comparable English text. This approach captures hand gestures on video and turns them into text that a non-signer can comprehend. Previously, similar research was conducted, with the bulk of them focused solely on sign translation of English alphabets or numerals. Hand gestures will be classified using the CNN algorithm. This technology will bridge the communication gap between signers and non-signers. This will make communication simpler for people who have difficulty speaking.

American Sign Language is the dominant sign language Since the only disability that D&M people have is the ability to communicate and they cannot use spoken language, their only way of communicating is through sign language. Communication is the process of exchanging thoughts and messages in different ways such as words, signals, behaviours and images. Deaf and hard of hearing persons (D&M) utilise their hands to create various gestures to communicate their ideas to others. Gestures are nonverbal communications that are communicated and are understood visually. Sign language refers to deaf and dumb people's nonverbal communication.

Sign language is a visual language and consists of 3 main components:

Finger Spelling	**Word Level Vocabulary**	**Non-Manual Features**
Used to spell words letter by letter.	Used for the majority of communication.	Facial expressions and tongue, mouth and body.

Table 1 Components of visual sign language

After recognizing the gesture our camera method does even translates the sign language predicted over there.

In our first Radar Approach the gestures we try to recognize are shown in the Fig. 1 below:

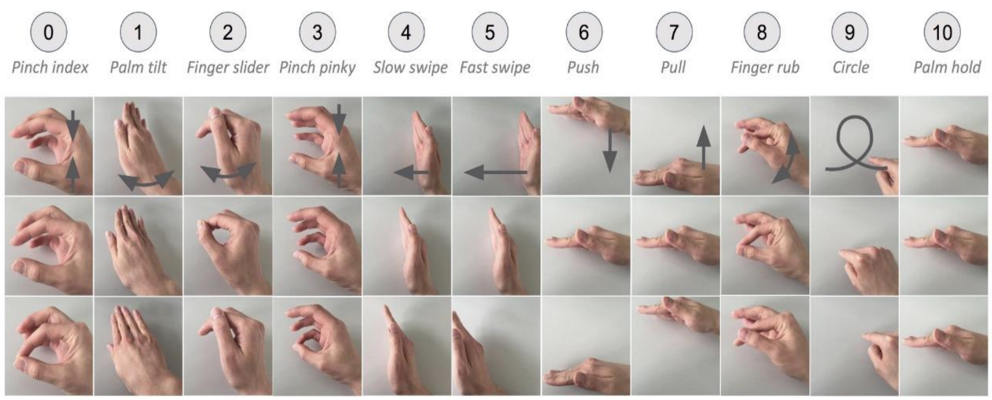


Figure 1: Deep Soli Gesture Data

In one of our approaches, we basically focused on recognizing finger-based hand gestures and creating a model that can be combined to form a complete word. The gesture you want to train is shown in the Figure 2 below:

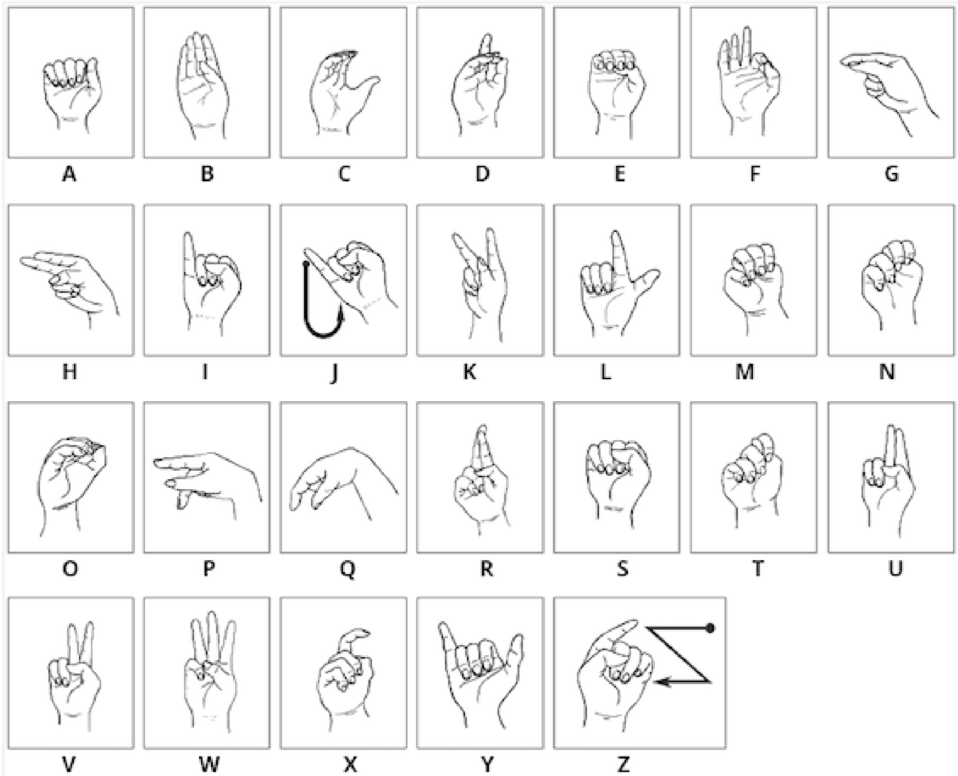


Figure 2: Sign conventions of alphabets in ASL

In our third approach with MediaPipe we'll try to classify the gestures like:

- okay
- peace
- thumbs up
- thumbs down
- call me
- stop
- rock
- live long
- fist
- smile

CHAPTER 2

LITERATURE SURVEY

In [1], their approach is a frequency-modulated continuous wave (FMCW) radar that can detect frequency shifts between transmitted and received electromagnetic waves according to the Doppler effect. The detection system is mounted on a radar array consisting of three continuous wave radars operating at 24.125GHz. A decision tree algorithm is constructed and evaluated as a classifier for the experiment. As a result, gesture movements are recorded according to the Doppler effect of the frequency spectrum of the radar signal. This system has higher in-plane motion detection performance than flexion and extension, and has achieved a high detection accuracy of 92% or more.

[2] Ultrasonic Frequency Modulated Continuous Wave (FMCW) and ConvLSTM models. One transmitter and three receivers spatially installed in different directions. The FMCW signal transmitted by the transmitter is reflected by hand and then detected by the receiver. Next, we obtained a range Doppler map (RDM) of the received signal by processing a 2D fast Fourier transform. High resolution at a distance of 0.005 m and speed of 0.03 m / s for hand gestures and 85.7% accuracy are achieved with the small size of 50 training samples of finger gestures.

In [3], we used the DCNN and VGG16 algorithms. ASL hand gesture movements are captured as microwave Doppler signals using a microwave X-band Doppler radar transceiver. These hand gestures are analyzed using MATLAB. The DCNN algorithm is used to train ASL gestures represented in spectrograms. The average validation accuracy of the DCNN and VGG16 algorithms was 87.5% and 95%, respectively.

In [4], we used FMCW radar and Deep Convolutional Neural Network (DCNN). The FMCW radar operated in the 24GHz ISM frequency band, the effective isotropic radiation power level was 0 dBm, and the FMCW radar received only one channel. Gesture recognition is performed using a deep convolutional neural network trained and tested in the Micro Doppler spectrogram. After training and validation, both methods yielded 99% classification accuracy in the test set.

In [5], we worked on FMCW radar, signal processing, and detection by a convolutional neural network (CNN). A sequence of object distance, velocity, and azimuth information was merged into a single input and sent to a convolutional neural network for learning spatial and temporal patterns. VGG10 converged to the 10th epoch with a verification accuracy of 92%. ResNet20 is superior to VGG10 with 98% verification accuracy. CNN + LSTM has the lowest accuracy in LEFT / RIGHT due to the lack of AngleOfArrival information, and the model achieves an average accuracy of 98% in the test set.

FasterRCNN, FMCW radar is used in [6]. A region-based deep complex neural network (RDCNN) is proposed to detect and classify gestures measured by a frequency modulated continuous wave radar system. In addition to the signature µD, we combine the phase difference information of the signal received from the L-shaped antenna array. The input of the proposed array contains three channels, i.e., one spectrogram and two out of phase channels. Results achieved 95% (96%) average PPV (APR) for nine gestures.

DopNet dataset the hand gesture radar dataset is used for gesture classification. The micro doppler signature is used as model input in [7]. Separate convolutional neural networks are used here. In order to accumulate without overloading, the paper uses a dissectible convolutional neural network model that performs deep convolution followed by point convolution. Get 94.56 accuracy on Dop Net data. Also, the computational time is minimized by using separable convolutions.

The Micro Doppler image dataset contains 15 types of sign language actions captured by radar echo as measured by the MDHandNet model in [8] The proposed MDHandNet model for hand gesture or sign language recognition. The accuracy obtained was 97.1%. Compared with other methods, the proposed model has good performance, fewer parameters and lower computational complexity.

Radar system, Soli dataset, Dop Net dataset using Spiking neural network have been deployed in [9]. The signal to collision conversion scheme is used to encode the Doppler radar map into spike trains fed to spike neural networks. The SNN's reader signal is fed into different classifiers.

The dataset 20BNjester (Open Source - Video Dataset) is used in combination between 3DCNN and LSTM (deep learning) networks in [10]. The proposed architecture extracts spatial-temporal information from input video sequences while avoiding many computations. 3DCNN is used to extract spatial and spectral features, which are then transmitted to the LSTM for classification.

The HBA system with infrared hand gesture image dataset contains 38,425 images in [11]. Convolutional Neural Network for Double Teacher Mode, a non-verbal approach to extracting features from infrared images of hand gestures. A robust nonlinear neural network is built with three convolutional layers. From the comparison experiments, the results demonstrate that the developed method can estimate hand gestures with a final accuracy of more than 90%.

In [12] localization and mapping algorithms are used. The images were first extracted from video streams recorded at different distances and locations. Each session consisted of

300 images from 6 different layers, each with 50 images. It describes the extraction of hand features by extracting the number of objects and the orientation in the object of interest. After doing the experiments, we can see that the overall accuracy of the system is 95.44%, which is a very good result. Here, good threshold values play an important role for gesture recognition.

Many researchers have proposed numerous methods for Hand Gesture Recognition Systems.

Electronics Based

- Use of electronic hardware
- Complex to use
- Physical attachment with system required
- Lot of noise in transition

Glove Based

- Requires use of hand gloves.
- Complex to use.
- Variable glove size for more users.
- Environment Dependent.

Marker Based

- Use of Color markers on finger or wrist.
- Marker positions are fixed.
- Complex to use multiple markers.

ISSUES TO BE ADDRESSED:

Latency

Image processing can be significantly slow creating unacceptable latency for videogames and other similar applications.

Robustness

Due to issues such as inadequate backdrop light, strong background noise, and so on, many gesture recognition systems can not recognize motions precisely or optimally.

Lack of Gesture Language

Different users make gestures differently, causing difficulty in identifying motions.

Performance

The image processing needed in gesture detection is highly resource expensive, making it difficult for apps to run on resource restricted devices.

CHAPTER 3

Survey Questions and Need Analysis Report

3.1 Survey Questions:

KALASALINGAM
Academy of Research and Education
DEEMED TO BE UNIVERSITY
Estd.U/S 3 of UGC Act 1956. Accredited by NAAC with "A" Grade

Anand Nagar, Krishnankoil - 626126. Srivilliputtur (Via), Virudhunagar (Dt), Tamil Nadu | info@kalasalingam.ac.in | www.kalasalingam.ac.in

CSE18R399 Community Service Project
EVEN 2021-22 Survey Questions

Title: AI Driven Hand Gesture Recognition

Survey Questions

S. No	Question	Response
Question 1:	Is There a Need for Hand Gesture Recognition Technology?	Yes No
Question 2:	Do you think it's unhygienic to use a touchscreen of a public device such as an informational table, especially in times of the pandemic?	Yes No
Question 3:	Nowadays, more and more people don't want to use touchscreens.	Strongly agree Agree Neutral Disagree Strongly disagree
Question 4:	For people like car drivers, touchscreens are often unsafe because the former need to distract their attention from driving to perform certain operations with the car's navigation system.	Strongly agree Agree Neutral Disagree Strongly disagree

Question 5:	**Touchscreens can be touched accidentally when you don't want it to.**	Yes No
Question 6:	**Touchless interfaces can solve these and many other problems touchscreens have.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 7:	**Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non-verbal form of communication.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 8:	**Touchless devices more convenient, hygienic, innovative and appealing, especially for millennials.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 9:	**The gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload.**	Yes No
Question 10:	**Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc.**	Strongly agree Agree Neutral Disagree Strongly disagree

Name:

Contact Number:

Occupation:

Address:

—..

Name Of the Students:

1. Marisetty Sai Murali (9919004342)
2. Yalla Sireesha (9919004304)
3. V Ramakrishna (9919004291)
4. Yalla Praveen (9919004303)

Signature Of the Guide:

R Raja Subramanian

Department of Computer Science and Engineering,

Kalasalingam Academy of Research and Education,

Virudhnagar, Tamil Nadu, India

Signature Of the concern target Community:

Name:

Place:

Designation:

Seal (If available):

3.2 NEED ANALYSIS

KALASALINGAM
Academy of Research and Education
DEEMED TO BE UNIVERSITY
Estd.U/S 3 of UGC Act 1956. Accredited by NAAC with "A" Grade
Anand Nagar, Krishnankoil - 626126. Srivilliputtur (Via), Virudhunagar (Dt), Tamil Nadu | info@kalasalingam.ac.in | www.kalasalingam.ac.in

CSE18R399 Community Service Project
EVEN 2021-22 Survey Questions

Title: AI Driven Hand Gesture Recognition

Survey Questions

S. No	Question	Response
Question 1:	Is There a Need for Hand Gesture Recognition Technology?	● Yes ○ No
Question 2:	Do you think it's unhygienic to use a touchscreen of a public device such as an informational table, especially in times of the pandemic?	● Yes ○ No
Question 3:	Nowadays, more and more people don't want to use touchscreens.	○ Strongly agree ○ Agree ● Neutral ○ Disagree ○ Strongly disagree
Question 4:	For people like car drivers, touchscreens are often unsafe because the former need to distract their attention from driving to perform certain operations with the car's navigation system.	○ Strongly agree ● Agree ○ Neutral ○ Disagree ○ Strongly disagree

Question 5:	**Touchscreens can be touched accidentally when you don't want it to.**	● Yes ○ No
Question 6:	**Touchless interfaces can solve these and many other problems touchscreens have.**	Strongly agree ■ Agree Neutral Disagree Strongly disagree
Question 7:	**Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non-verbal form of communication.**	■ Strongly agree Agree Neutral Disagree Strongly disagree
Question 8:	**Touchless devices more convenient, hygienic, innovative and appealing, especially for millennials.**	Strongly agree ■ Agree Neutral Disagree Strongly disagree
Question 9:	**The gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload.**	● Yes ○ No
Question 10:	**Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc.**	■ Strongly agree Agree Neutral Disagree Strongly disagree

Name: Dhana Lakshmi

Contact Number: 9390039616

Occupation: Software Engineer

Address: Hyderabad

Dhanalakshmi

Name Of the Students:

1. Marisetty Sai Murali (9919004342)
2. Yalla Sireesha (9919004304)
3. V Ramakrishna (9919004291)
4. Yalla Praveen (9919004303)

Signature Of the Guide:

R Raja Subramanian

Department of Computer Science and Engineering,

Kalasalingam Academy of Research and Education,

Virudhnagar, Tamil Nadu, India

Signature Of the concern target Community:

Name: Dhana Lakshmi

Place: Hyderabad

Designation:

Software Engineer

Seal (If available):

CSE18R399 Community Service Project
EVEN 2021-22 Survey Questions

Title: AI Driven Hand Gesture Recognition

Survey Questions

S. No	Question	Response
Question 1:	**Is There a Need for Hand Gesture Recognition Technology?**	Yes No
Question 2:	**Do you think it's unhygienic to use a touchscreen of a public device such as an informational table, especially in times of the pandemic?**	Yes No
Question 3:	**Nowadays, more and more people don't want to use touchscreens.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 4:	**For people like car drivers, touchscreens are often unsafe because the former need to distract their attention from driving to perform certain operations with the car's navigation system.**	Strongly agree Agree Neutral Disagree Strongly disagree

Question 5:	**Touchscreens can be touched accidentally when you don't want it to.**	Yes No
Question 6:	**Touchless interfaces can solve these and many other problems touchscreens have.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 7:	**Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non-verbal form of communication.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 8:	**Touchless devices more convenient, hygienic, innovative and appealing, especially for millennials.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 9:	**The gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload.**	Yes No
Question 10:	**Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc.**	Strongly agree Agree Neutral Disagree Strongly disagree

Name: SK. Basha

Contact Number: 7075480895

Occupation: Teacher

Address: Vijayawada

SK Basha

Name Of the Students:

1. Marisetty Sai Murali (9919004342)
2. Yalla Sireesha (9919004304)
3. V Ramakrishna (9919004291)
4. Yalla Praveen (9919004303)

Signature Of the Guide:

R Raja Subramanian

Department of Computer Science and Engineering,

Kalasalingam Academy of Research and Education,

Virudhnagar, Tamil Nadu, India

Signature Of the concern target Community:

Name: Sk. Basha

Place: Vijayawada

Designation:

Teacher

Seal (If available):

KALASALINGAM
Academy of Research and Education
DEEMED TO BE UNIVERSITY
Estd.U/S 3 of UGC Act 1956. Accredited by NAAC with "A" Grade

Anand Nagar, Krishnankoil - 626126. Srivilliputtur (Via), Virudhunagar (Dt), Tamil Nadu | info@kalasalingam.ac.in | www.kalasalingam.ac.in

CSE18R399 Community Service Project
EVEN 2021-22 Survey Questions

Title: AI Driven Hand Gesture Recognition

Survey Questions

S. No	Question	Response
Question 1:	**Is There a Need for Hand Gesture Recognition Technology?**	Yes (selected) No
Question 2:	**Do you think it's unhygienic to use a touchscreen of a public device such as an informational table, especially in times of the pandemic?**	Yes (selected) No
Question 3:	**Nowadays, more and more people don't want to use touchscreens.**	Strongly agree Agree (selected) Neutral Disagree Strongly disagree
Question 4:	**For people like car drivers, touchscreens are often unsafe because the former need to distract their attention from driving to perform certain operations with the car's navigation system.**	Strongly agree (selected) Agree Neutral Disagree Strongly disagree

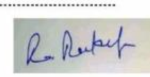
Question 5:	**Touchscreens can be touched accidentally when you don't want it to.**	Yes No
Question 6:	**Touchless interfaces can solve these and many other problems touchscreens have.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 7:	**Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non-verbal form of communication.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 8:	**Touchless devices more convenient, hygienic, innovative and appealing, especially for millennials.**	Strongly agree Agree Neutral Disagree Strongly disagree
Question 9:	**The gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload.**	Yes No
Question 10:	**Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc.**	Strongly agree Agree Neutral Disagree Strongly disagree

Name: R. Rakesh

Contact Number: 9032426089

Occupation: Software Engineer

Address: Chennai



Name Of the Students:

1. Marisetty Sai Murali (9919004342)
2. Yalla Sireesha (9919004304)
3. V Ramakrishna (9919004291)
4. Yalla Praveen (9919004303)

Signature Of the Guide:

R Raja Subramanian

Department of Computer Science and Engineering,

Kalasalingam Academy of Research and Education,

Virudhnagar, Tamil Nadu, India

Signature Of the concern target Community:

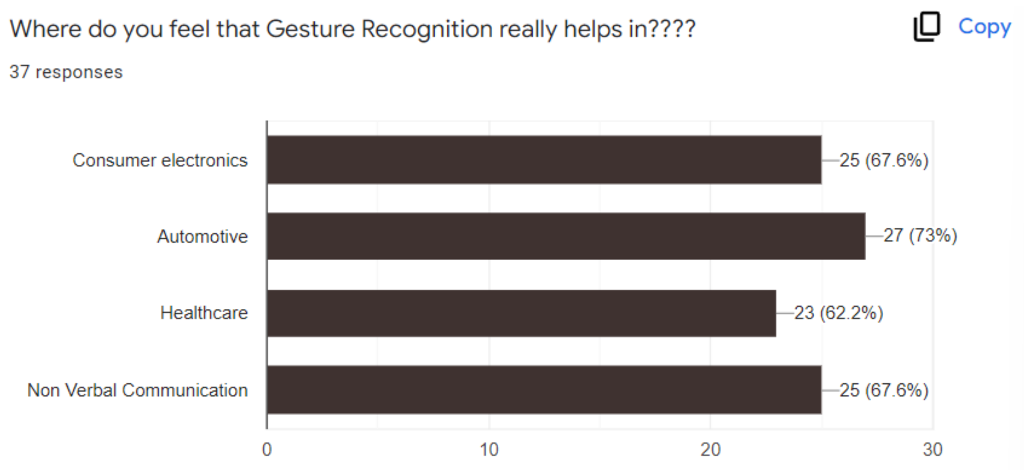
Name: R. Rakesh

Place: Chennai

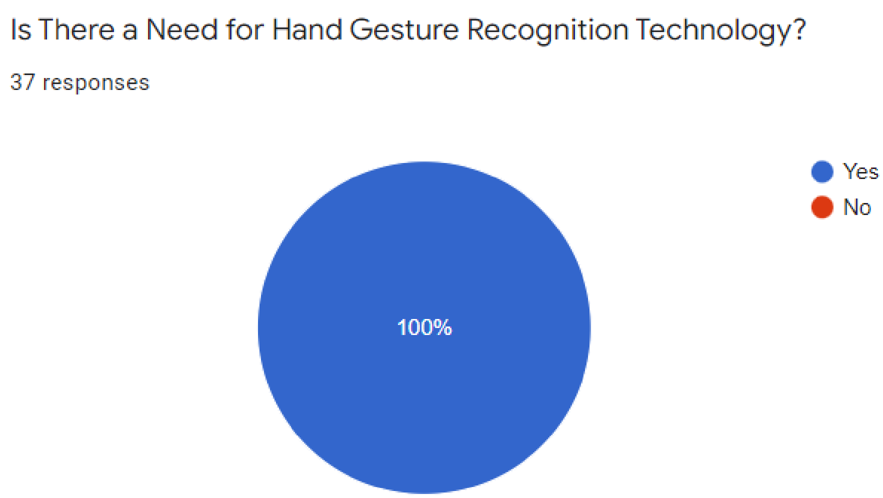
Designation:

Software Engineer

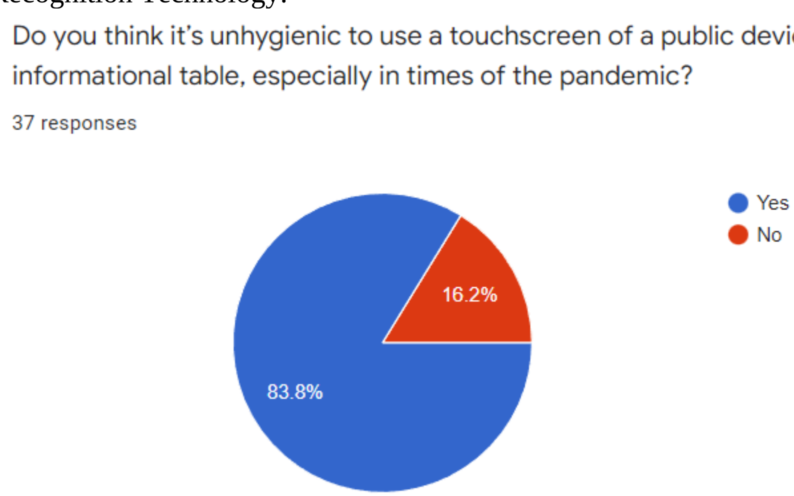
Seal (If available):



- Most of the people felt that the gesture recognition relly helps in Automative, Non-verbal Communication, Consumer Electronics and then Health Care.



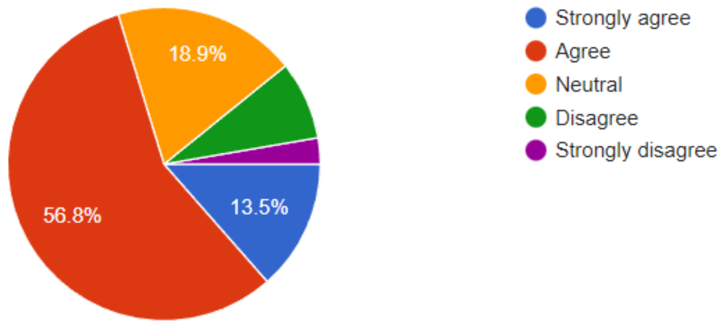
- All the people who took our survey believe that there's a 100% need of Hand Gesture Recognition Technology.



- When we asked if you think it's unhygienic to use a touchscreen of a public device such as an informational table, especially in times of the pandemic 84% people felts it's true.

Nowadays, more and more people don't want to use touchscreens.

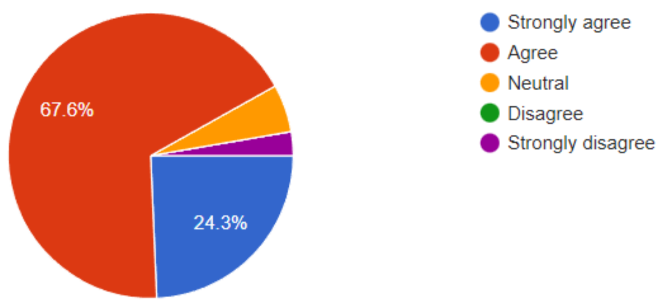
37 responses



- Nowadays, more and more people don't want to use touchscreens.

For people like car drivers, touchscreens are often unsafe because the former need to distract their attention from driving to perform certain operations with the car's navigation system.

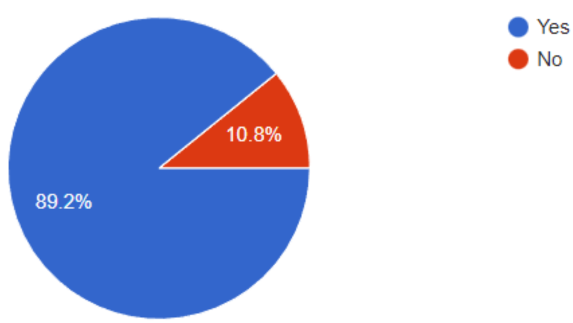
37 responses



- People agreed to the point that for people like car drivers, touchscreens are often unsafe because the former need to distract their attention from driving to perform certain operations with the car's navigation system.

Touchscreens can be touched accidentally when you don't want it to.

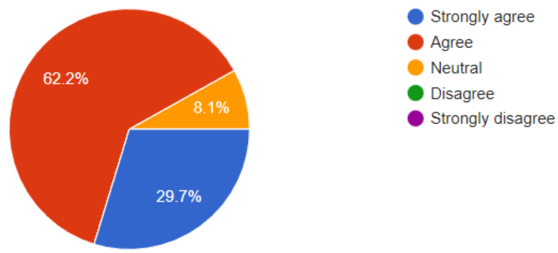
37 responses



- Around 90% people believe that touchscreens can be touched accidentally when we don’t want it to.

Touchless interfaces can solve these and many other problems touchscreens have.

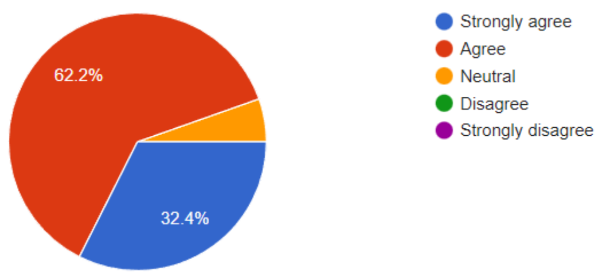
37 responses



- Survey analysis on Touchless interfaces can solve these and many other problems touchscreens have can be seen clearly in the pie chart above.

Hand gestures are the natural way of interactions when one person is communicating with one another and therefore hand movements can be treated as a non-verbal form of communication.

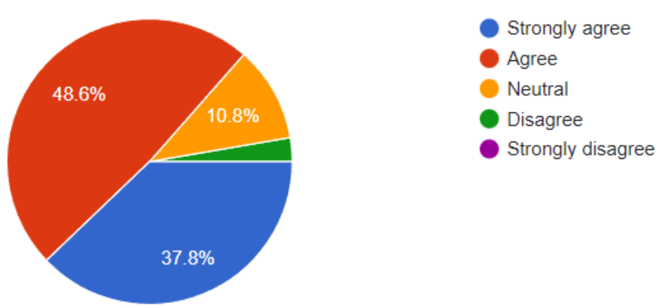
37 responses



- Hand gestures are the natural manner of exchanges when one person is interacting with another, and so hand motions may be viewed as a nonverbal form of communication, according to 63 percent of individuals.

Touchless devices more convenient, hygienic, innovative and appealing, especially for millennials.

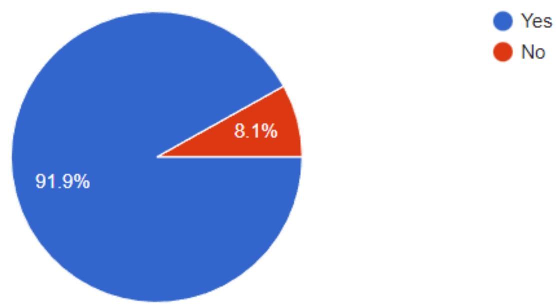
37 responses



- There was a mixed response for the above question.

The gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload.

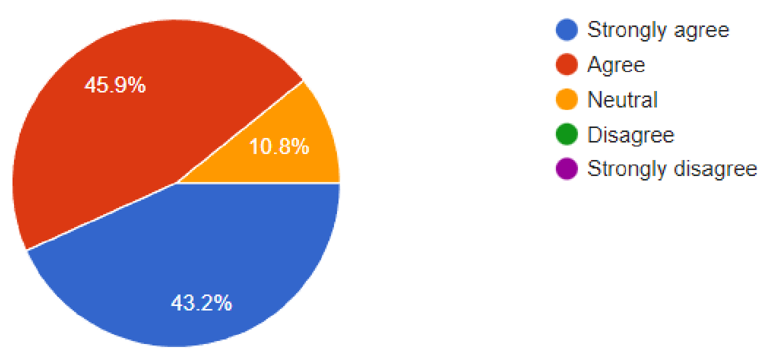
37 responses



- 90% were clear that the gestural interaction is a promising means to cover the full range of driver's operational needs while minimizing the visual workload.

Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc.

37 responses



- Response for Hand gesture recognition is of great importance for human computer interaction (HCI) because of its extensive applications in virtual reality and sign language recognition etc. is pictured in the above pie chart.

CHAPTER 4

OBJECTIVES

The main objective of this project is to look for new ways to combine traditional radar signal processing techniques with Data-Driven approaches to provide a better hand gesture recognition model.

- Use different sensors towards hand gesture recognition.
- Create a gesture recognition model that can help
- Compare which approach is better like Sensor vs. Camera vs. Camera with MediaPipe.
- Conduct a literature review on hand gesture recognition using radar sensors, machine learning and deep learning.
- Create a sign language recognition system that can help the people who are deaf and hard of hearing to communicate.

CHAPTER 5

METHODOLOGY

In Hand gesture Recognitio feature extraction and representation is representing an image as a 3D matrix with dimensions such as the height and width of the image and the value of each pixel as depth (1 for grayscale, 3 for RGB).
In addition, these pixel values are used to extract useful features using CNN.

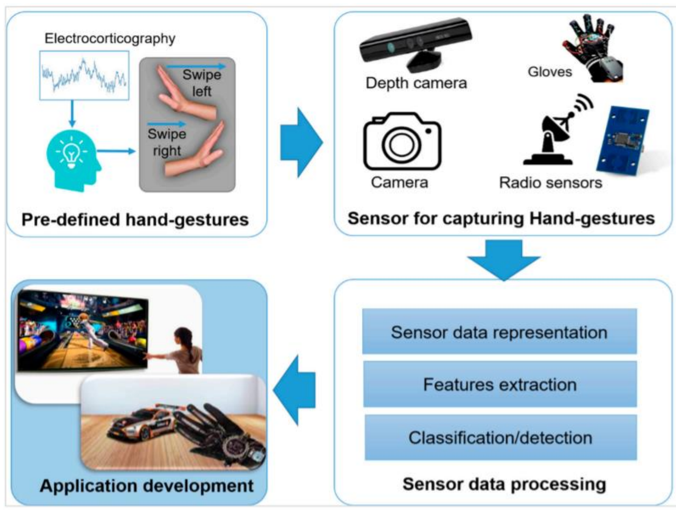


Figure 3: Hand gesture recognition architecture.

Note:
I learnt the concept of Classification in the course CSE18R212 - Machine Learning

5.1 Artificial Neural Networks

Artificial neural networks are connections of neurons that mimic the structure of the human brain. Each connection in a neuron sends information to another neuron. Input is supplied to the first layer of the neuron, which processes them and sends them to another layer of the neuron called the hidden layer. After processing the information through multiple layers of the hidden layer, the information is passed to the final output layer.

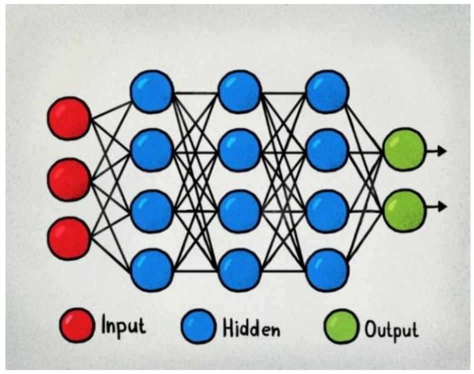


Figure 4: Artificial Neural Network

5.1.1 Unsupervised Learning:

Unsupervised learning is a type of machine learning that looks for previously undetected patterns in an existing unlabeled data set and with minimal human supervision. The two most important unsupervised learning methods are principal component analysis and cluster analysis.

The only requirement, called unsupervised learning strategy, is to learn a new feature space that captures the properties of the original space by maximizing the objective function or minimizing the loss function. Therefore, the covariance matrix generation is not applicable. Even though the learning is unsupervised, we obtain the eigensymbols of the covariance matrix because the eigen analysis operations of linear algebra maximize the variance. This is called principal components analysis.

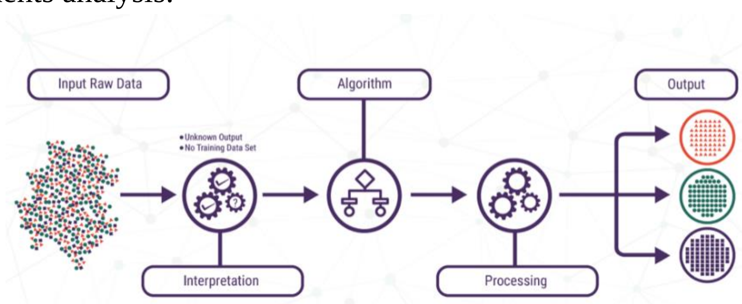


Figure 5: Unsupervised Learning

5.1.2 Supervised Learning:

Supervised learning is a machine learning task that learns a function that maps an input to an output, e.g., an input/output pair. It derives a function from labeled training data consisting of a series of training examples. In supervised learning, each example is a pair of input objects (usually vectors) and desired output values (also known as screens). Supervised learning algorithms analyze the training data and generate derived functions that can be used to map new examples. The best-case scenario allows the algorithm to correctly determine the class labels for the hidden cases. This requires generalizing the learning algorithm "reasonably" to situations that are not visible from the training data.

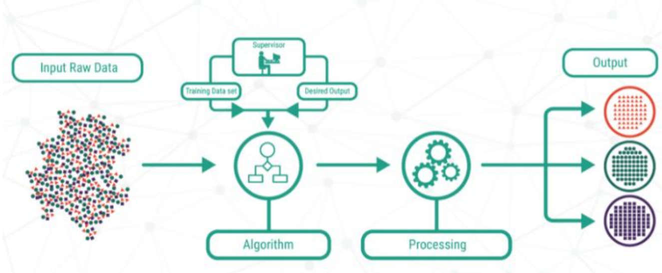


Figure 6: Supervised Learning

5.1.3 Reinforcement Learning:

Reinforcement learning (RL) is an area of machine learning. It deals with how software agents behave in their environment to maximize the concept of cumulative reward. Reinforcement learning, along with supervised and unsupervised learning, is one of the three basic paradigms of machine learning. Reinforcement learning differs from supervised learning in that it does not require the presentation of labeled input/output pairs and the need to explicitly modify suboptimal actions. Instead, the focus is on finding a balance between exploration (uncharted territory) and mining (current knowledge).

Figure 7: Reinforcement Learning

5.2 Convolution Neural Networks:

Unlike a regular neural network, the neurons in the CNN layer are arranged in three dimensions: width, height, and depth. The neurons in a layer are connected to only a small region of the layer (window size). Before that, instead of all fully connected neurons. Also, at the end of the CNN architecture, the final output layer has dimensions (number of layers) to reduce the overview to a single vector of layer scores.

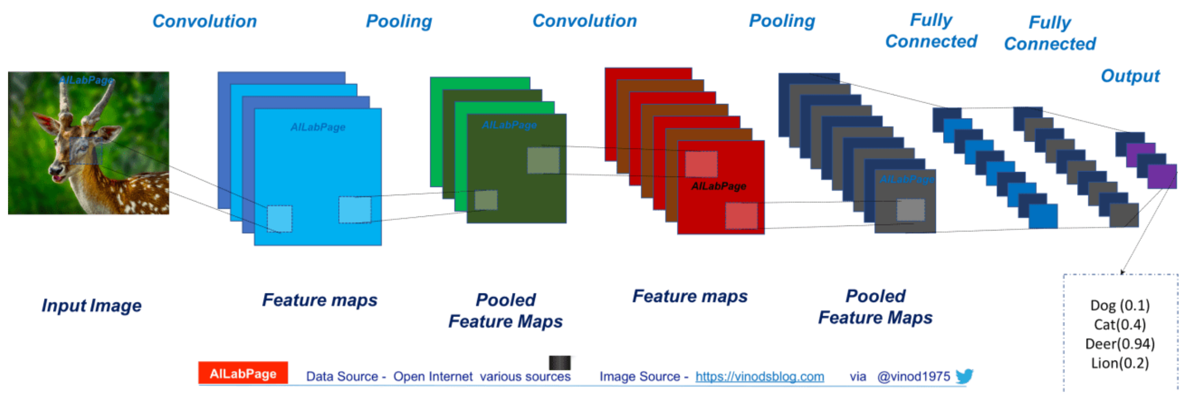


Figure 8: Convolutional Neural Network

5.2.1. Convolution Layer:

The convolution layer uses a small window size (typically 5 * 5 length) that extends to the depth of the input matrix. This layer consists of window-sized learnable filters. During each iteration, we moved the window incrementally [usually 1] to calculate the product of the filter entry and the input value at a particular position. Continue this process to create a two-dimensional activation matrix that reflects the response of this matrix at each spatial location. This means the network will learn

A filter that is activated when a visual feature is displayed. B. Edges in a particular direction or spots of a particular color.

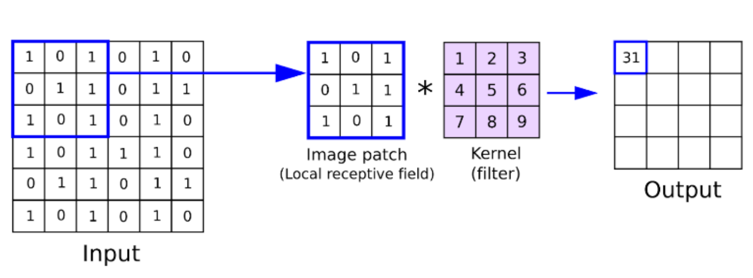


Figure 9: Convolutional Layer

5.2.2 Pooling layer:

Use the pooling layer to reduce the size of the activation matrix and eventually reduce the parameters that can be learned there.

There are two types of pooling.

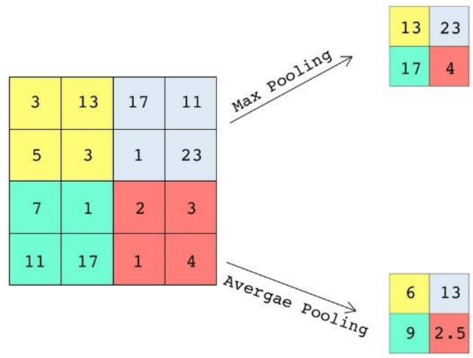


Figure 10: Maximum and Average Pooling

5.2.2.a Maximum pooling:

Maximum pooling uses the window size [Example: Size 2 * 2 window example] And just take up to 4 value. If you close this window well and continue this process, you will end up with an activation matrix that is half the original size.

5.2.2.b Average pooling:

Average pooling takes the average of all the values in the window.

5.2.3 Fully connected layers:

In the convolutional layer, neurons are only connected to the local area, but in a fully connected area, all inputs are properly connected to the neuron.

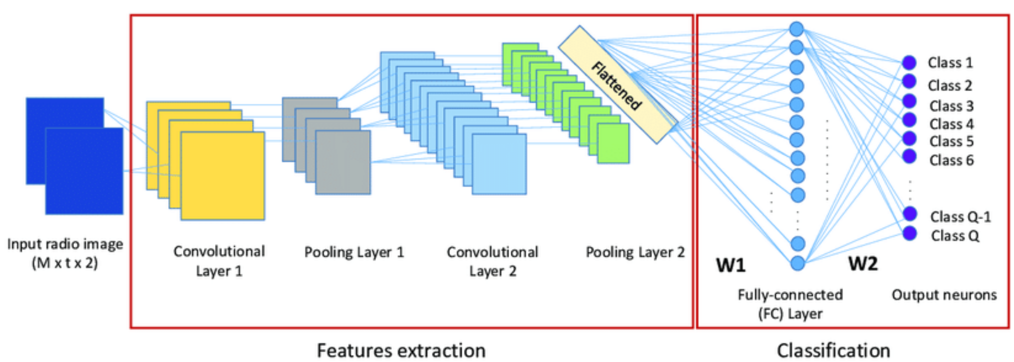


Figure 11: CNN Architecture

5.2.4 Final output layer:

After getting the values from the fully connected layers, connect them appropriately to the last layer of the neuron (a number equal to the total number of classes). This predicts the probability that each image will be in a different class.

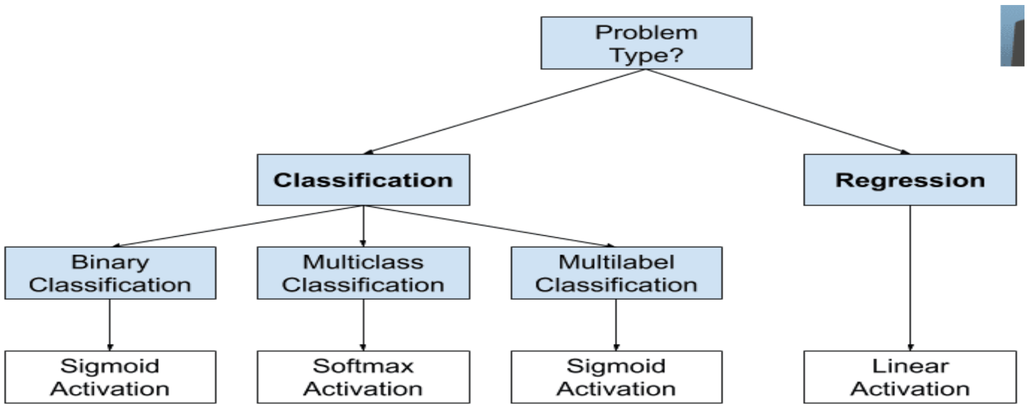


Figure 12: Activation Functions for different output layers

5.3 TensorFlow

TensorFlow is an open-source numerical software library calculation. First define the nodes of the calculation graph, then the actual calculation is done within the session. TensorFlow is widely used in machine learning.

5.4 Keras

Keras is a high level of neural network library written in Python, and acts as a wrapper for TensorFlow. This is, to build a neural network with a minimum of code lines, is used when you want to quickly build and test. Layer, the objective function, activation function, optimizer, etc., contains the implementation of the neural network elements that are commonly used. Since the tool is easy to operate the image and text data.

5.5 OpenCV

OpenCV (Open-Source Computer Vision) is an open-source library of programming functions used for real-time computer vision. It is mainly used for analysis of functions such as image processing, video recording, and face and object recognition. Written in C ++, the primary interface, Python, Java, and MATLAB / OCTAVE bindings are available.

CHAPTER 6

RESULTS AND DISCUSSION

Approach 1:

6.1. Radar Dataset:

6.1.1 Google Deep Soli Radar Gesture Dataset:

The dataset contains multiple pre-processed Range-Doppler Image sequences. Each sequence is saved as a single HDF5 format data file. File names are defined as [gesture ID] _ [session ID] _ [instance ID].h5. Range-Doppler Image data of a specific channel can be accessed by dataset name ch[channel ID]. Label can be accessed by dataset name label. Range-Doppler Image data array has shape of [number of frames] * 1024 (can be reshape back to 2D Range-Doppler Image to 32 * 32)
Dataset session arrangement for evaluation.

- 11 (gestures) * 25 (instances) * 10 (users) for cross user evaluation: session 2 (25), 3 (25), 5 (25), 6 (25), 8 (25), 9 (25), 10 (25), 11 (25), 12 (25), 13 (25).
- 11 (gestures) * (50 (instances) * 4 (sessions) + 25 (instances) * 2 (sessions)) for single user cross session evaluation: session 0 (50), 1 (50), 4 (50), 7 (50), 13 (25), 14 (25).
- Please refer to the paper for the gesture collecting campaign details.

The gestures are listed in the table below. Each column represents a gesture, and we take a snapshot of three important steps for each gesture. The gesture label is represented by the number in the circle above. Please note that the order of the gesture labels is different from the order on the paper, because we grouped the gestures in the paper. Strings with Gesture ID 11 are background cues without the hand's face.

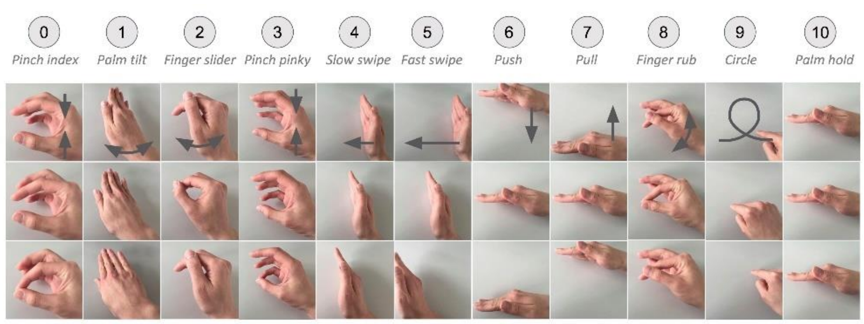


Figure 13: Google Deep Soli Gesture Radar Data

The Soli sensor is a solid-state millimeter wave radar. Conventional radar approaches rely on high spatial resolution to distinguish many rigid moving targets (e.g., aircraft). In contrast, Soli uses high temporal resolution priority detection to detect subtle, non-rigid movements. Soli uses a single, wide antenna beam to illuminate the entire hand as modulated pulses are transmitted at a very high repetition rate (between 110 kHz).

The received raw signal, consisting of superposition of reflections from scattering centers in the radar antenna beam, is then processed into some abstract representation of the signal. The high temporal resolution allows a combination of fast time and slow processing to map the scattering central reflection into interpretable dimensions. Range Doppler images (RDI) are obtained by mapping the received energy in two-dimensional space to radial distance (or range) and velocity. Pixel intensity is reflected energy; horizontal axis is velocity; vertical axis is range.

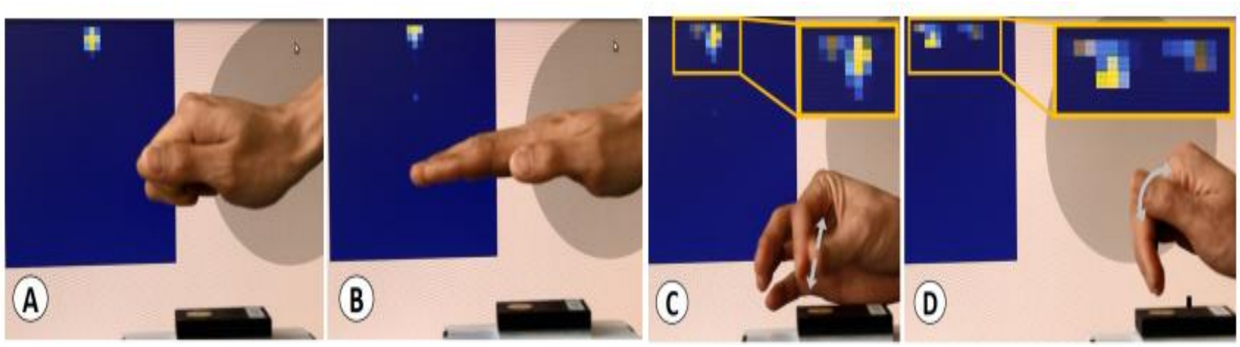


Figure 14: Testing the Radar with static and dynamic gestures.

(A+B) Sensor produces almost identical response for static objects even of distinct shape

Short-range radar data does not directly contain information about shape and hence many existing algorithms are not applicable.

The steps to process the data is:

Signal transformations (e.g., Fourier transform).
Feature Extraction.
Frame-level classification.

Each data point is a sequence of variable length of 1024 dimensional vector which are reshaped to a [32,32] 2d Doppler map [range, velocity].

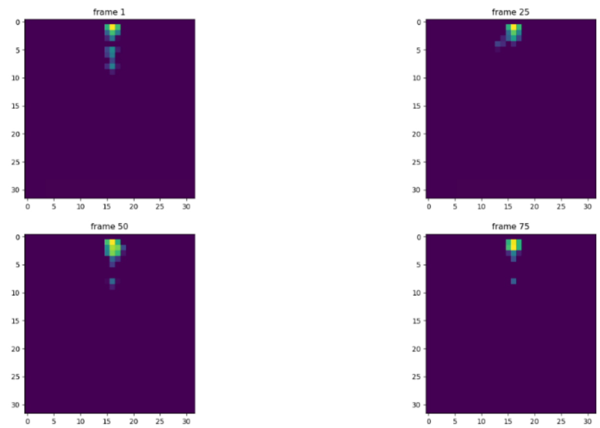


Figure 15: Radar Gesture Intensity.

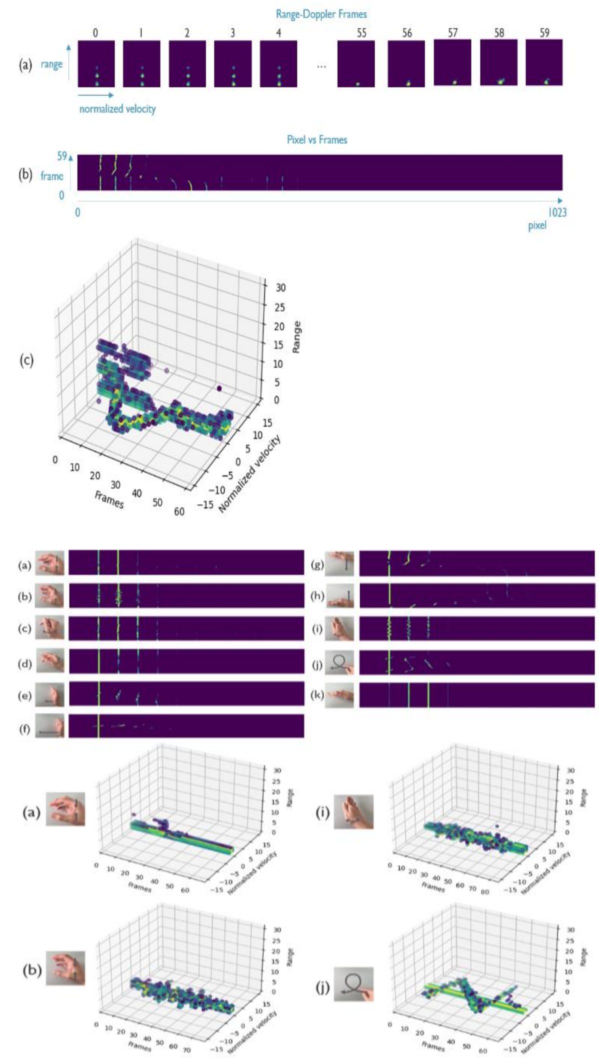


Figure 16: Visualising Radar Data.

Approach 2:

6.2. Camera Sensor Data:

Image Dataset acquired from : https://github.com/loicmarie/sign-language-alphabet-recognizer/tree/master/dataset

6.2.1: Dataset:

We've got the data set from the images dataset given above, the dataset have images of letters from A-Z and then Del, Nothing and Space. For now, we'll be training the model with only 10 alphabets and we'll try to translate the sentence too.

First, we capture each frame shown by the webcam of our machine. In each frame we define a region of interest (ROI) which is denoted by a blue bounded square as shown in the image below.

Figure 17: RGB Image Of Alphabet "E".

From this whole image we extract our Region of Interest (ROI) which is RGB and background of the Image as shown below.

Figure 18: Image Of Alphabet "E" after removing Background Noice.

Finally, we apply our mask filter to our image which helps us extracting various features of our image. The image after applying mask blur looks like below.

Figure 19: Image of Alphabet "E" after applying mask

MediaPipe Approach:

6.3 MediaPipe Hand Gesture Recognition:

Gesture recognition is an active research area in human-computer interaction engineering. It has many uses such as controlling virtual environments, translating sign language, controlling robots and producing music. This hand gesture recognition machine learning project uses OpenCV and Python's MediaPipe framework and TensorFlow to build real-time hand gesture recognition.

OpenCV is a real-time C/C-based computer vision and image processing framework. However, it is used in Python via the Python OpenCV package.

6.3.1: MediaPipe:

MediaPipe is a customizable machine learning solution framework developed by Google. It is an open source and extremely lightweight cross-platform framework. MediaPipe comes with pre-trained ML solutions like face detection, pose estimation, hand detection, object detection and more.

6.3.2 Tensor Flow:

TensorFlow is an open-source machine learning and deep learning library developed by the Google Brains team. It can be used for a variety of tasks, but with a particular focus on deep neural networks.

Neural networks are also known as artificial neural networks. It is a subset of machine learning and is the heart of deep learning algorithms. The concept of neural networks is inspired by the human brain. It mimics the way biological neurons signal each other. A neural network consists of a node layer that includes an input layer, one or more hidden layers, and an output layer.

We'll first use MediaPipe to recognize the hand and the hand key points. MediaPipe returns a total of 21 key points for each detected hand.

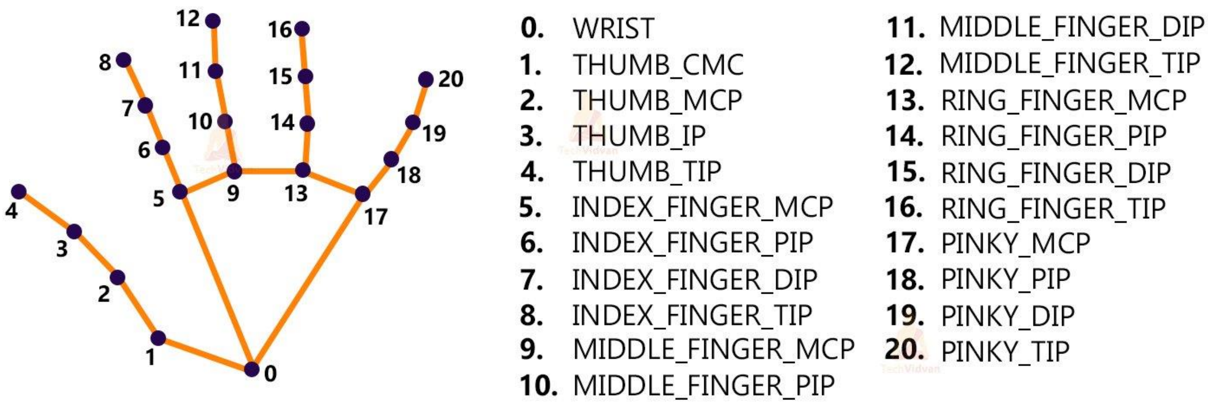


Figure 20: Key points detected by MediaPipe

CHAPTER 7

CONCLUSION AND FUTURE WORK

AI Driven Hand Gesture Recognition through different approaches. We tried different approaches towards AI Driven Hand Gesture Recognition. First, we tried Radar Gesture Recognition with Google Deep Soli Radar Gesture Dataset. Though we achieved good accuracy over there, the radar data failed to recognize the static gestures and then the Radar has it advantages as it can survive in any climate and it can penetrate through some mediums too. Sine, the Radars were unable to recognize the Static Gestures and Radar Data need lot of computational power. Now, we tried camera approach where we had to recognize the American Sign Language, we trained out 10 digits and then we have included a translator too in the program and then we had good accuracy over there too but we had around 60% validation accuracy over there. Since, it is very time conserving as we had to train a lot of gestures and then we have to preprocess the images, the captured images then go through a filter to remove background and then a mask to extract the main features and then the images will be gone through a CNN to classify the gestures over there and then the sentence can be translated using text to speech. The third approach is using the MediaPipe, the MediaPipe recognizes the key points of the hand and then the trained models will be used to detect the hand gestures over there.

Due to the disadvantages, we faced while running the model, we believe that adding images of all classes with different backgrounds and distances will cover a wider area and flexibility for live demonstrations. To further improve the model, you need to add a mirrored image as part of the left-handed dataset to perform data expansion to make it translation-invariant / equivariant. This prevents the model from over-adjusting certain backgrounds, distances, and hand positions in the image. This allows you to merge the current model with the new dataset to improve the prediction accuracy of each character. Overall, the more images with different characteristics, the better the results.

Instead of just recognizing letters, we can try to extend the model to recognize words, phrases, and coherent phrases. Once we have developed a model for generating text/annotations that form logical sentences, we can continue with natural language processing so that we can run various analyzes based on the text we extracted. output, such as sentiment analysis to find context. and the feelings behind the words. This is significant because the model will combine two different cognitive computing methods: image recognition and text analysis. In other words, this model will become one of the representative examples of how to extract text from images.

Publication:

R. Raja Subramanian, Marisetty Sai Murali, Vedala Ramakrishna, Yalla Sireesha, Yalla Satya Praveen, "AI Driven Hand Gesture Recognition", Elsevier Procedia, 2022 (Submitted)

REFERENCES

[1] Lan, Shengchang; He, Zonglong; Yao, Kai; Chen, Weichu (2018). *[IEEE 2018 IEEE/MTT-S International Microwave Symposium - IMS 2018 - Philadelphia, PA, USA (2018.6.10-2018.6.15)] 2018 IEEE/MTT-S International Microwave Symposium - IMS - Hand Gesture Recognition using a Three-dimensional 24 GHz Radar Array. , (), 138–140.* doi:10.1109/MWSYM.2018.8439658

[2] Fanchen Kong, Jingcheng Deng, Zichuan Fan, Gesture recognition system based on ultrasonic FMCW and ConvLSTM model, Measurement, Volume 190, 2022, 110743, ISSN 0263-2241.

[3] H. Kulhandjian, P. Sharma, M. Kulhandjian and C. D'Amours, "Sign Language Gesture Recognition Using Doppler Radar and Deep Learning," 2019 IEEE Globecom Workshops (GC Wkshps), 2019, pp. 1-6, doi: 10.1109/GCWkshps45667.2019.9024607.